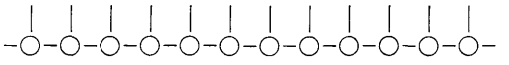
A SCALED MODEL OF A PROPOSED PROTEIN STRUCTURE[1]

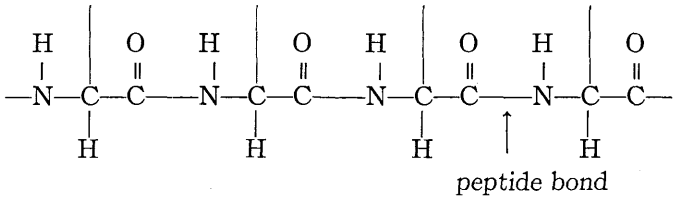
EDWARD MACK, JR.
University of North Carolina and Battelle Memorial Institute

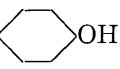
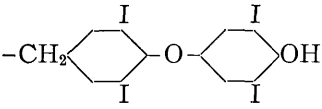
This little article is intended to give a brief view of our present quite imperfect picture of protein molecule structure, and is addressed particularly to some of the geometrical aspects of the situation.

In a certain sense it is possible that the structure of the protein molecules is very simple. According to the classical theory, based largely on Emil Fischer's work, a protein molecule is a chain-like structure, the nature of which may be indicated in the following manner:



where the members of the chain are amino acid residues. The small circles, or heads, are the same for all amino acids (we shall pass over the imino acids). They are joined together by the "peptide" bond, formed by splitting off water between the $-COOH$ group in the head of one amino acid and the $-NH_2$ group in the head of a neighboring amino acid, thus:



The tails, or side chains, are different for different amino acids. For glycine, the side chain is $-H$; for alanine $-CH_3$; for lysine $-CH_2CH_2CH_2CH_2NH_2$; for tyrosine  ; for thyroxine  ; etc. About twenty-four different amino acids seem to be important in animal metabolism, although a few others are known.

[1]The model described here was first presented at a Symposium at the Richmond Meeting of the American Association for the Advancement of Science in December, 1938; and again at the Columbus Meeting of the Ohio Academy of Science in May, 1940.

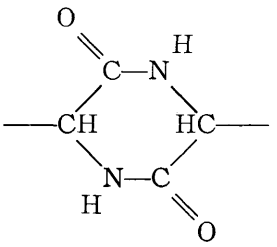
In another sense of the word, protein structure may be considered as not at all *simple*. The average molecular weight of the amino acid is about 120. Since the molecular weight of various proteins range from perhaps about 34,000 up to 100,000,000 or more, it is obvious that a very large number of amino acid residues are present in such giant molecules. It has been suggested that the many shades of physical and chemical properties, which the proteins are known to possess, are to be accounted for not only by the actual total number of amino acid residues present, and by the particular kinds of residues present, and by their relative population but also by the nature of their sequence along the molecule backbone. A chain –GGGAAAGGGAAA–, where G and A represent glycine and alanine, might well produce an entirely different set of behaviors than the arrangement –GAGAGAGAGAGA–. With only a few different amino acids in a backbone which is merely a few hundred links long, the number of possible permutations is, of course, immensely large. It seems likely, too, that a regularly repeated order of arrangement would lead to quite different properties than a random one.

Bergmann (1) and others have, in fact, shown that the amino acids are probably arranged in some definite order, in natural protein molecules, and in some cases it has now been possible to deduce the sequence. "Living organisms therefore do not achieve the synthesis of the immense number of proteins provided by the peptide theory in its original conception, but seem to synthesize only those proteins that exhibit—simple numerical rules. The discovery of the quantitative rules governing the protein molecule makes the biological synthesis of an individual protein molecule appear as a process which involves a specificity that is both highly delicate and extremely complex. Such specificity phenomena are a clear indication of the operation of an enzymatic process which directs all the steps involved in the synthesis of the special pattern of each individual protein."

STRING, SHEET AND BODY STRUCTURES

There is reasonably good evidence for the conclusion that some protein molecules exist as strings, others as sheets or as 3-dimensional lumps of some sort. It is not unlikely that a given molecule may assume all of these shapes in different settings. Long ago, for example, Dr. Nouy (2) estimated the size and shape of certain protein molecules lying in a water surface, and concluded that they were tiny cylinders or prisms, probably formed by the coiling up of the long string-like molecules. Mirsky and Pauling (3) have pointed out that the well-defined properties of native proteins require that their molecules have definite configurations. Whether such protein molecules, lying alongside one another, are joined together laterally by chemical bonds (somewhat after the manner of vulcanized rubber molecules), or by hydrogen bonds, or by van der Waals forces, strong or weak, is a question to which we would like to have an unambiguous answer.

In this connection, an extremely interesting theory has been advanced by Dr. Wrinch (4)—the cyclol fabric. It is supposed that two amino acid molecules unite to form a closed diazine ring, by splitting off water as in the formation of the peptide bond:



The formation of such a hexagonal ring is plausible. Indeed, if the two side chains are H atoms, this diazine ring is diketopiperazine, which is known to exist. Corey (5), by X-ray examination of the crystal, has found that the diketopiperazine rings are flat, and that they are joined together by hydrogen bonds in long strings.

Now, according to the early Wrinch theory, these diazine rings unite chemically with one another by the reaction

$$>C=O+H-N<\longrightarrow>\overset{\overset{OH}{|}}{C}-N<$$

to give a structure, in which three diazine rings meet in a triazine ring, and which by extension in space yields the so-called cyclol fabric (Figure 1) a sheet-like structure. Further, it is supposed that the fabric can fold into a "cage," which is a truncated regular tetrahedron. Pictures of the model of this structure are shown in some of Dr. Wrinch's papers, but will not be reproduced here. The smallest cage of this type contains 72 amino acids, and the larger ones $72x2^2$, $72x3^2$, $72x4^2$, etc. Since the average molecular weight of amino acids is about 120, the molecular weights of these protein cage molecules (forgetting the smallest) would be about 34,000, 76,500, 136,000, etc. This seems to be in remarkably close accord with the actually observed non-random distribution of molecular weights; although an alternative explanation which has been offered is that chain molecules of certain definite lengths, with none in between, may be formed because of the nature of the act of protein synthesis.

SCALED MODEL OF CYCLOL FABRIC

While, as already stated, the formation of the diazine rings is plausible, a very serious difficulty confronts the hypothesis that these rings can join together to form the fabric. The difficulty can be shown with the aid of a scaled model (Figure 2).

The central carbon atom of the heads of the amino acids, with the exception of glycine, are all assymetric, and happen to be of the l-variety. As a direct consequence of this, the nests of three OH groups on the alternate triazine rings are located above and below the plane of the fabric (see Figure 1), whereas the side chains are all above the plane. Half of the side chains are disposed "initially upwards," the other half "initially outwards" as shown in Dr. Wrinch's diagram (Figure 1) and correspondingly shown in Figure 2 by the match sticks (which represent the side chains).

The model is made on a scale of 10^8 to 1, or 1 cm.≡A°. The C–C internuclear distance is taken as 1.54, C–N as 1.47, C–H as 1.08, C–O as 1.43, and O–H as 0.97. The diameter of the hydrogen atom (more acceptably set at 2.13 or larger) is shaded down, for the sake of the argument, to 2.0; that is, the domain radius of the hydrogen atoms, beyond the nucleus, is made to be 1.0. This is in substantial accord with the value estimated (6) from the crystal lattices of methane and ethane,

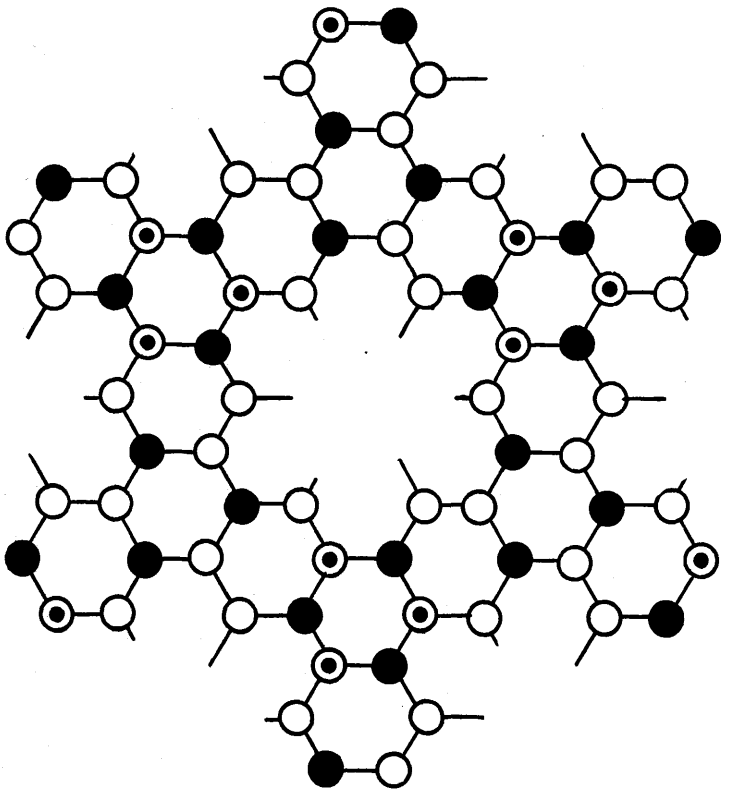


Fig. 1. A fragment of the cyclol fabric. Solid symbols represent nitrogen. Hollow symbols represent C(OH), hydroxyl upwards. Hollow symbols with dot in center represent C(OH), hydroxyl downwards. Hollow symbols with long arm represent CHR, direction of side chain initially outwards. Hollow symbols with short arm represent CHR, direction of side chain initially upwards.

and with the size needed to account for the cross-sectional areas of hydrocarbon chains in crystals and in oil films. The angle of the oxygen bonds is taken as 111°, the carbon bonds are tetrahedral, and I have allowed myself the license of making the angles of the three bonds of the nitrogen atom 109½°. (The angles are probably more like 108°). This license permits the assembly of the model as a "puckered" planar sheet, and amounts really to only a very small concession in the direc-

tion of conforming with Dr. Wrinch's practice. As a matter of fact, as far as the conclusion regarding the plausibility of the cyclol fabric is concerned, one can increase or decrease the bond angles by many degrees, without affecting the final clear cut conclusion.

One difficulty lies in the region of the fabric indicated by the white arrow in Figure 2. The three hydroxyl groups (shown here with oxygen atoms to scale) find sufficient room on the triazine rings for their accom-

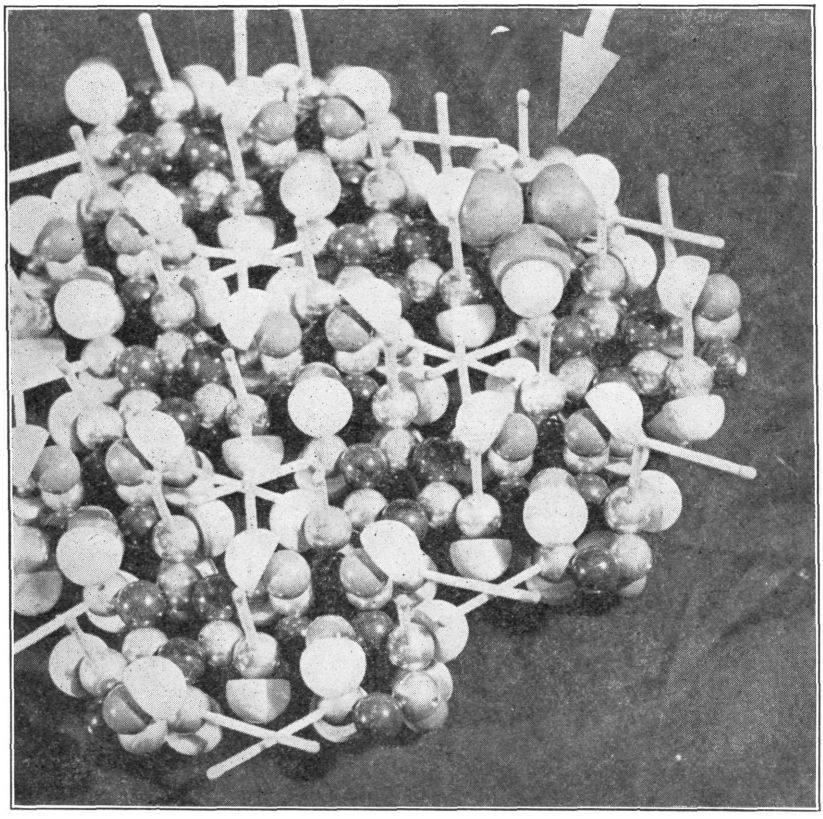


Fig. 2. View of model from above.

modation, either as formal OH groups, or with the H atoms resonating between pairs of O atoms. But the three "initially upward" side chains surrounding the rest of hydroxyls are very severely crowded for space.

A far more serious difficulty involves the "initially outward" side chains. Three of these extend into the central hexagonal cavity of Figure 1, and are shown as three crossed match sticks in the corresponding cavities of Figure 2. Now, it is to be borne in mind that these match sticks represent side chains which, on the model scale, are at least as bulky (with the exception of glycine) as normal hydrocarbon chains,

and in some of the amino acids they are as bulky as benzene rings (tyrosine and thyroxine, for example). It is obviously simply physically impossible to pack so much matter into so small a space.

Figure 3, which is Figure 1 reproduced more nearly to proper atomic scale, shows the difficulty clearly. HHH are the three hydrogen atoms which are carried by the same carbon atoms to which the three vertical ("initially upward") side chains, around the rim of the cavity, are also

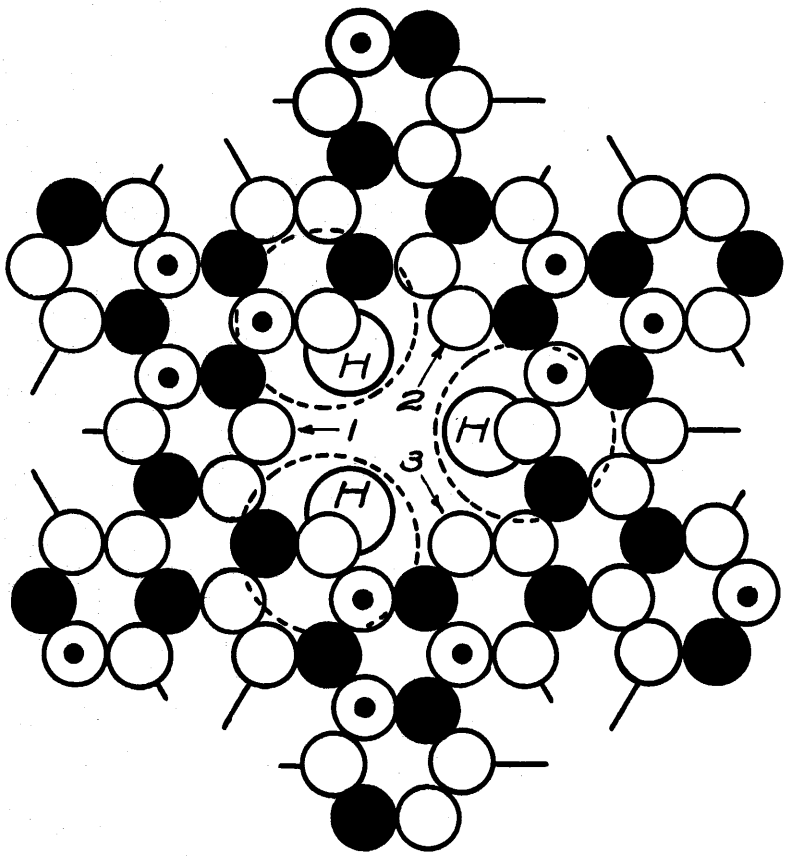


Fig. 3. Fragment of cyclol fabric drawn roughly to scale to show extremely small relative size of hexagonal cavity.

attached. The large dotted circles show the minimum space occupied by these vertical side chains of about 18 sq. A°. The points where the side chains "initially outwards" are supposedly attached are indicated by 1, 2, 3. The free space left in the hexagonal cavity, into which these three "initially outward" side chains are supposed to protrude, from their points of attachment to three carbon atoms bordering the cavity, and out of which they must extend and reach up (or down) to the surface in some way, is barely large enough to accommodate a single

extra hydrogen atom. Only if the three hydrogen atoms HHH were entirely removed *and* only if the three *vertically* placed side chains were entirely removed, would the free space of the cavity take care of a *single* normal hydrocarbon chain; and many of the amino acid side chains are larger in cross-section than a hydrocarbon chain, and furthermore *three* side chains are to be accommodated.

The protein fibroin is said to contain glycine to the extent of about 50% of its amino acid residues (1). It is conceivable that the Wrinch fabric could account satisfactorily for it and other possible proteins of at least 50% glycine content, *if* all of the glycine hydrogens were to occupy the "initially outward" positions, and *if* the "initially upward" side chains could also be attached to the fabric. The difficulty is that most proteins seem to contain much less than 50% glycine residues and many contain none at all.

Huggins (7) has already called attention to these space difficulties, as have also Neurath and Bull (8). Pauling and Niemann (9) have shown the improbability of the formation of the cyclol fabric on the grounds of energy considerations.

Hydrogen bond linkage of the diazine rings has been suggested (10), and more recently Dr. Wrinch (11) has described a modified fabric in which instead of a lactam-lactim there is a keto-enol transformation. But still the available space is far from adequate to permit the proposed disposition of the side chains. Huggins (12) has also made some further suggestions, which deserve consideration.

It is with regret that one feels compelled to abandon the cyclol fabric and its close cousins. Many of the attractive features of the cyclol fabric theory have been set forth by Langmuir (13). Dr. Wrinch's speculations have been immensely stimulating to those interested in protein structure, and it is still possible that her general picture of protein structure is not far from the truth. The objections raised here are directed against the specific cyclol fabrics which have been suggested.

On the other hand, we should not forget that a satisfactory accounting for the facts of protein behavior may very likely be rendered by a development and extension of the classical theory of long chain peptide structure; and as illustration of such an extension the recent fascinating and suggestive paper of Pauling (14) on "A Theory of the Structure and Process of Formation of Antibodies" may be cited.

REFERENCES

(1) **Bergmann.** Chem. Rev., **22,** 423 (1938).
(2) **Du Nouy.** Phil. Mag., **48,** 264 (1924).
(3) **Minsky and Pauling.** Proc. Nat. Acad. Sci., **22,** 439 (1936).
(4) **Wrinch.** Nature, **137,** 411 (1936); Proc. Roy. Soc. London, **160,** 59 (1937); etc.
(5) **Corey.** J. Am. Chem. Soc., **60,** 1598 (1938).
(6) **Mack.** J. Am. Chem. Soc., **54,** 2141 (1932).
(7) **Huggins.** J. Am. Chem. Soc., **61,** 755 (1939).
(8) **Neurath and Bull.** Chem. Rev., **23,** 391 (1938).
(9) **Pauling and Niemann.** J. Am. Chem. Soc., **61,** 1860 (1939).
(10) **Wrinch and Lloyd.** Nature, **138,** 758 (1936).
(11) **Wrinch.** Phil. Mag., **30,** 64 (1940).
(12) **Huggins.** J. Chem. Phys., **8,** 598 (1940).
(13) **Langmuir.** Proc. Phys. Soc. London, July (1939).
(14) **Pauling.** J. Am. Chem. Soc., **62,** 2643 (1940).